

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Costly information acquisition and the temporal resolution of uncertainty

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1689031> since 2019-02-04T11:51:23Z

Published version:

DOI:10.1016/j.jmateco.2015.07.005

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Costly information acquisition and the temporal resolution of uncertainty

Daniele Pennesi¹

*THÉMA - University of Cergy-Pontoise,
33 Boulevard du Port, Cergy-Pontoise.*

Abstract

This paper studies the choice of an individual who acquires information before choosing an action from a set of actions, whose consequences depend on the realization of a state of nature. Information processing can be costly, for example, due to limited attention. We show that the preference of the individual is completely characterized by a preference for early resolution of uncertainty, which becomes indifference when facing degenerate choices. When information acquisition is no longer part of the decision process, the individual is indifferent to the timing of resolution of uncertainty and she behaves according to the subjective learning model of [Dillenberger et al. \(2014\)](#).

Keywords: Costly Information Acquisition, Menu Choice, Rational Inattention, Timing of Resolution of Uncertainty

JEL CLASSIFICATION: D01, D9, D8

1. Introduction

Information, uncertainty and time are the essential dimensions of most economic decisions. Typically, information is used to reduce uncertainty and uncertainty resolves over time. In many situations, however, acquiring information may be costly: for example, when individuals have bounded computational abilities. The present paper studies the interaction of time, uncertainty and costly information acquisition.

¹Email: daniele.pennesi@u-cergy.fr - The research was supported by Labex MMEDII (ANR-11-LBX-0023-01). I thank Mark Machina and three anonymous referees for helpful suggestions.

Consider an American tourist who is willing to visit one European country, either Greece (G) or Italy (I). Imagine them as "menus of actions". For example the actions in G may be {Kos, Athens} and the actions in I may be {Florence, Venice, Rome}. The payoff of each action depends on the state of nature that will realize: the weather. The choice of the destination, G or I , depends on the exchange rate between the dollar and the Euro.² The tourist wants to acquire information about the weather forecasts to select the best option from each menu and she may allocate attention to different web's sites. She can do it in two different ways

1. Before observing the exchange rate
2. After observing the exchange rate

In the first case, she will acquire information concerning the weather forecasts of both Greece and Italy and she will schedule her trip contingently to the exchange rate. If she observes the exchange rate before acquiring information, she will pay attention only to the weather forecast that is relevant to the location she will visit. It is reasonable to imagine that an individual who pays a cost to acquire information will prefer to observe the exchange rate before bearing that cost. The result of the paper establishes that this is essentially *the only consideration of an individual who pays a cost to acquire information*. In other words, we interpret uncertainty resolving before information acquisition (as in 2.), as early resolution of uncertainty. Whereas, late resolution corresponds to a randomization device (the exchange rate) that resolves its uncertainty after information is acquired. We show that the preference over menus of actions of an individual who acquires costly information, is characterized by a preference for early resolution of uncertainty, which becomes indifference when the decision is degenerate. This last consideration follows from the interpretation of the information acquisition problem: information is instrumental to perform a better choice *from* the menu, when the menu contains a single action there is no choice to make.

In addition, when the individual is indifferent toward the timing of resolution of uncertainty, i.e. indifference between observing the exchange rate before or after scheduling her vacation, her preferences are represented by the Subjective Learning model of [Dillenberger et al. \(2014\)](#).

²For example, if the rate is higher than a given threshold, the tourist will choose G , otherwise she will choose I . Let assume that, with probability α , the rate is higher than the threshold.

With respect to the standing literature, the current paper identifies the conditions concerning preferences for the timing of resolution of uncertainty, which are necessary and sufficient to characterize Costly Information Acquisition (CIA). Indeed, as shown in Ex. 1, an alternative property, often interpreted as representing "preference for early resolution of uncertainty", may be too weak to characterize CIA.

The results of the paper raise questions concerning observational distinguishability of a model of costly information acquisition from a pure preference for the timing of resolution of uncertainty. As pointed out above, information is instrumental to select the best action from a menu of actions. When facing degenerate menus, i.e. singletons, costly information acquisition is not effective. Therefore, the timing of resolution of uncertainty for choices from degenerate menus is irrelevant. This implies that CIA is distinguishable from a pure preference for early resolution of uncertainty. This distinction can be used to experimentally discriminate the CIA model from those valuing early resolution of uncertainty intrinsically.

Concerning the Subjective Learning model, there is no observational distinguishability from a pure indifference toward the timing of resolution of uncertainty.

2. Overview of the results

An individual with CIA preferences optimally chooses an act f from a set of acts F . Each act associates a payoff $f(\omega) \in X$ to the state of the world $\omega \in \Omega$ that will realize. Information about the true state of the world is acquired through a *signal* (or experiment) at a cost c . The signal specifies the probability of forming a posterior p given a prior \hat{p} . After information is received, the decision maker selects an act from the menu. The problem can be formulated as:

$$\max_{\pi \in \Pi(\hat{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega)) p(d\omega) \right) \pi(dp) - c(\pi) \right] \quad (\text{CIA})$$

where $u : X \rightarrow \mathbb{R}$ is a utility over the payoffs in X , $\Pi(\hat{p})$ is the set of all signals for a given prior \hat{p} and $c : \Pi(\hat{p}) \rightarrow [0, \infty]$ is the cost of information.

The CIA model has been axiomatized by [De Oliveira et al. \(2014\)](#) and it is interpreted as representing the preferences of a Rationally Inattentive individual (in the sense of [Sims, 2003](#)). The timing of the rational inattentive choice procedure is the following: an unobservable (to the decision maker) state of the world is selected by nature, the decision maker chooses a menu from which

she will select an act later. After choosing a menu, information is acquired and the posterior is formed. Lastly, the individual selects an act from the menu and she receives the payoff associated with the state of the world and the selected act. The decision maker may also face randomized menus (as in the exchange rate example), $\alpha F + (1 - \alpha)G$. In this case, she chooses a contingent plan $f \in F$ and $g \in G$, and a randomization device determines which of them will be implemented. $f \in F$ with probability α and $g \in G$ with probability $(1 - \alpha)$. For randomized menus there are two layers of uncertainty: the state of the world and the realization of the randomization device. In the axiomatization of [De Oliveira et al. \(2014\)](#), the realization of the randomization device is known to the individual only *after information is acquired*. Hence, the individual always sets up a contingent plan of actions. In the present work, we assume a more

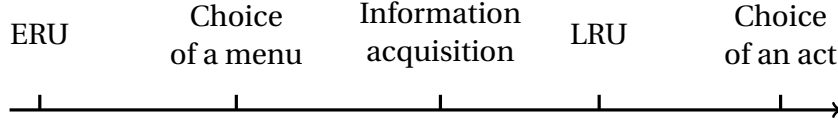


Figure 1: Early (ERU) and late (LRU) resolution of uncertainty

general domain in which revealed preferences are observed. Our setting allows to distinguish between later and earlier resolution of uncertainty. Indeed, our primitive is a preference over lotteries over menus of acts. We allow for a different form of randomization between menus that resolves its uncertainty *before* information is acquired. Figure 1 illustrates the timing of the decision process in the two cases. LRU is the point in time at which randomization takes place in [De Oliveira et al. \(2014\)](#), we allow an additional form of randomization that resolves its uncertainty at ERU.

As in [Ergin and Sarver \(2015\)](#), the early randomization³ is modeled through the lottery $\alpha\delta_F \oplus (1 - \alpha)\delta_G$. With probability α , it pays a degenerate lottery delivering F , with probability $1 - \alpha$ it pays a degenerate lottery delivering G and it is played immediately. It corresponds to the left-hand side of Fig. 2. In this case, the individual knows the menu from which she will choose in the second stage, before acquiring information.

Differently, when facing $\delta_{\alpha F + (1 - \alpha)G}$ the individual acquires information to

³The mixture operator \oplus represents mixture of lotteries, whereas the symbol $+$ mixture of menus (see Section 3).

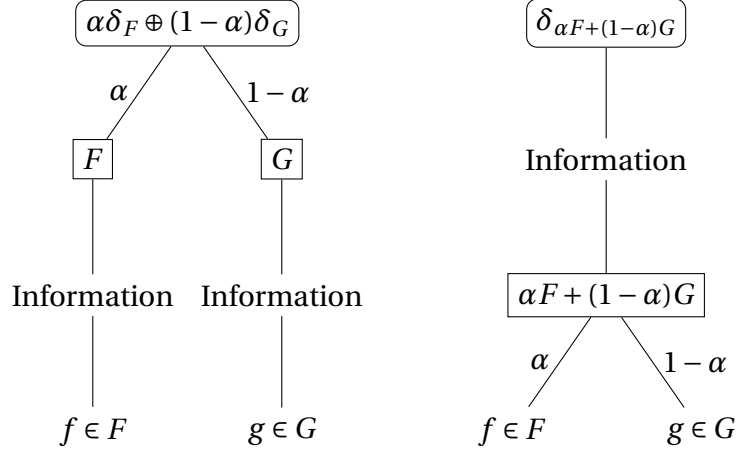


Figure 2: Early resolution of uncertainty (left) and late resolution of uncertainty (right).

perform contingent choices from menus and the realization of the randomization device determines which plan will be implemented.

It is easy to see that a CIA individual will prefer the left-hand side to the right-hand side of Fig. 2. This is due to the cost of acquiring information, since the choice in the right-hand side of Figure 2 is "more complex" in terms of the amount of information to analyze, hence more costly.

The opposite implication, that preferring the left-hand side to the right-hand side of Figure 2 implies CIA is, however, not true in general. As claimed above, when facing randomization between degenerate menus, the acquisition of information becomes superfluous, since information is used to perform an optimal choice from the menu. Therefore, denoting by f the degenerate menu $\{f\}$, the CIA model implies $\alpha\delta_f \oplus (1-\alpha)\delta_g \sim \delta_{\alpha f + (1-\alpha)g}$. This is not necessarily true for a pure preference for early resolution of uncertainty: in that case, an intrinsic preference for early resolution of uncertainty may follow from alternative reasons such as, anxiety or the possibility to take hidden actions (see [Ergin and Sarver \(2015\)](#)), hence a strict preference $\alpha\delta_f \oplus (1-\alpha)\delta_g > \delta_{\alpha f + (1-\alpha)g}$ is possible. Theorem 1 establishes that the CIA model is completely characterized by the existence of a strict preference for early resolution of uncertainty when non-degenerate menus are taken into account.⁴

⁴For some $F, G \in \mathcal{A}$ with $|F| > 1$ and $|G| > 1$ and $\alpha \in (0, 1)$, $\alpha\delta_F \oplus (1-\alpha)\delta_G > \delta_{\alpha F + (1-\alpha)G}$

When indifference toward the timing of resolution of uncertainty is extended to all menus, $\alpha\delta_F \oplus (1-\alpha)\delta_G \sim \delta_{\alpha F + (1-\alpha)G}$, preferences are represented by a particular case of CIA, namely the Subjective Learning (SL) model of [Dillenberger et al. \(2014\)](#). In the SL model, information is exogenously fixed, therefore it is not part of the decision problem. In the initial example, the American tourist is indifferent between knowing the exchange rate before or after acquiring information, if she can only gather information concerning the weather forecasts of a single country, for example Greece.

As a final result, we prove that the attitude toward the timing of resolution of uncertainty is related to the solution of the optimal information acquisition problem. Suppose there exists a source of information (a signal) π^* that solves the maximization in Eq. (CIA) for two menus F and G . If this is the case, the individual is indifferent between acquiring information before or after the randomization between F and G , i.e. $\alpha\delta_F \oplus (1-\alpha)\delta_G \sim \delta_{\alpha F + (1-\alpha)G}$. In other words, she "locally" behaves according to the SL model, since she can always acquire π^* .

2.1. Related literature

The acquisition of information may be costly for a variety of reasons. One of these, limited attention, has been recently studied by different authors. It was introduced with a series of papers by C. Sims ([Sims, 1998, 2003](#)) as a natural explanation for price stickiness. The model has been subsequently applied to a variety of topics.⁵

The behavioral foundation of the rational inattention model has been studied over the last years. Recently, [Caplin and Dean \(2015\)](#) and [Ellis \(2013\)](#) characterize rational inattentive preferences observing *ex post* choices from menus of alternatives. [Masatlioglu et al. \(2012\)](#) studied how attention is revealed from choice. [Ergin and Sarver \(2010a\)](#) studied the problem of costly information acquisition when the uncertainty is completely subjective, i.e. uncertainty about future tastes. Recently, [Piermont et al. \(2015\)](#) studied responsive learning, where an individual dynamically learns about her tastes through consumption. Differently from rational inattention, consumption affects information acquisition hence, the cost of information is represented by foregone consumption.

⁵Price setting ([Mackowiak and Wiederholt, 2009](#)), optimal consumption/saving ([Luo, 2008; Tutino, 2008](#)), business cycles ([Kacperczyk et al., 2009](#)), portfolio under-diversification ([Van Nieuwerburgh and Veldkamp, 2009](#)), asset pricing ([Mondria, 2010](#)), stochastic choice ([Matějka and McKay, 2015](#)).

In the current paper, the primitive is a preference over lotteries over menus of acts and we do not assume observability of choices from menus. The closest paper to the current one is [De Oliveira et al. \(2014\)](#), although their primitive is a preference over menus of actions. The focus of the current paper is to provide an alternative behavioral foundation of the same model, relying on the distinction between early and late resolution of uncertainty. A full comparison with [De Oliveira et al. \(2014\)](#) is discussed in Section 3.2. [De Oliveira \(2014\)](#) characterizes a particular case of the CIA model, with an entropic cost of attention that corresponds to the formulation of the RI model originally proposed by [Sims \(2003\)](#).

[Dillenberger et al. \(2014\)](#) axiomatized another particular case of the CIA model, called Subjective Learning (SL), where information is exogenously fixed. In this work we give an alternative axiomatization of the SL model.

From the technical viewpoint, this paper modifies the approach of [De Oliveira et al. \(2014\)](#) and [Dillenberger et al. \(2014\)](#) introducing an explicit distinction between early and late resolution of uncertainty, in the spirit of [Kreps and Porteus \(1978\)](#). The current paper is also related to the work of [Ergin and Sarver \(2015\)](#). They study intrinsic preference for early resolution of uncertainty when choices are observed over lotteries over menus of lotteries. In their model (second-stage) uncertainty is subjective since there are no "objective" states of the world. A full comparison with their work is contained in Section 3.3.

3. Axioms

A finite set Ω contains the states of the world, when a state is realized, all uncertainty is resolved. We denote by X the set of consequences, it is a convex subset of a topological vector space. An act is a function $f : \Omega \rightarrow X$, the set of all acts is denoted by \mathcal{F} . A constant act $f(\omega) = x$ for all $\omega \in \Omega$ is identified with an element of X . We denote by \mathcal{A} the set of all nonempty and finite subsets of \mathcal{F} . Given a topological space M , we denote by $\Delta(M)$ the set of probabilities defined on the Borel σ -algebra of M , endowed with the weak* topology. The binary relation \succsim representing preferences is defined over lotteries over menus of acts i.e. on $\Delta(\mathcal{A})$. \oplus denotes the standard mixture of two lotteries. We denote by f the singleton menu $F = \{f\}$. x denotes the singleton menu containing only a constant act $f(\omega) = x$ for all $\omega \in \Omega$ and some $x \in X$. δ_F denotes the degenerate lottery paying $F \in \mathcal{A}$ for sure. The mixture $+$ of two acts, $\alpha f + (1 - \alpha)g$ for all $\alpha \in [0, 1]$ is performed state-wise, $(\alpha f + (1 - \alpha)g)(\omega) = \alpha f(\omega) + (1 - \alpha)g(\omega)$. The

mixture of two menus $\alpha F + (1 - \alpha)G$ for all $\alpha \in [0, 1]$ is the menu of all mixtures of elements in F and G , $\alpha F + (1 - \alpha)G = \{\alpha f + (1 - \alpha)g : f \in F, g \in G\}$.

The first axiom includes some standard properties whose interpretation is well established.

Axiom (Preference).

1. **(Weak Order)** \succsim is a non-trivial weak order.
2. **(Continuity)** For all $P, Q, R \in \Delta(\mathcal{A})$, the sets $\{\alpha \in [0, 1] : \alpha P \oplus (1 - \alpha)Q \succsim R\}$ and $\{\alpha \in [0, 1] : R \succsim \alpha P \oplus (1 - \alpha)Q\}$ are closed.
3. **(Dominance)** For all menus $F \in \mathcal{A}$ and acts $g \in \mathcal{F}$, $\delta_{F \cup g} \succsim \delta_F$, with $\delta_{F \cup g} \sim \delta_F$ if there is $f \in F$ such that $\delta_{f(\omega)} \succ \delta_{g(\omega)}$ for each $\omega \in \Omega$.
4. **(Unboundedness)** There are $x, y \in X$ with $x \succ y$, such that for all $\alpha \in (0, 1)$, there exist $z, z' \in X$ such that $\delta_{\alpha z + (1 - \alpha)y} \succ \delta_x \succ \delta_{\alpha z' + (1 - \alpha)y}$.

Weak Order and Continuity are minimal rationality and continuity requirements. Dominance is introduced (in their setting) by [De Oliveira et al. \(2014\)](#), it implies a preference for flexibility, i.e. $G \subseteq F$ implies $\delta_F \succsim \delta_G$, and the monotonicity axiom used in choice under ambiguity i.e. $\delta_{f(\omega)} \succ \delta_{g(\omega)}$ for all $\omega \in \Omega$ implies $\delta_f \succ \delta_g$. Unboundedness implies that $u(X) = \mathbb{R}$, it only restricts preferences over final prizes.⁶

The next axiom is the classical von Neumann-Morgenstern independence:

Axiom (Ex-ante Independence). For any $P, Q, R \in \Delta(\mathcal{A})$ and $\alpha \in (0, 1)$,

$$P \succsim Q \iff \alpha P \oplus (1 - \alpha)R \succsim \alpha Q \oplus (1 - \alpha)R$$

The interpretation is standard, a preference for a lottery P over Q is not reversed when mixing with a third lottery.

The next definition formally introduces the preference for early resolution of uncertainty (PERU), (see [Ergin and Sarver \(2015\)](#)).

⁶Unboundedness is a technical requirement and it is a fairly common assumption (see [Cerrei-Vioglio et al. \(2011b\)](#); [De Oliveira et al. \(2014\)](#)). It implies the existence of arbitrarily good and arbitrarily bad outcomes. For example, if X is the set of simple lotteries over monetary prizes, unboundedness is satisfied when the individual's risk attitude is represented by a logarithmic utility. [De Oliveira et al. \(2014\)](#) assume unboundedness above or below, we assume unboundedness above and below.

(PERU). For any $F, G \in \mathcal{A}$ and $\alpha \in (0, 1)$,

$$\alpha \delta_F \oplus (1 - \alpha) \delta_G \succsim \delta_{\alpha F + (1 - \alpha) G}$$

It implies that the left "tree" of Figure 2 is weakly preferred to the right one.

Similarly, we define indifference toward the timing of resolution of uncertainty:

(ITRU). For any $F, G \in \mathcal{A}$ and $\alpha \in (0, 1)$,

$$\alpha \delta_F \oplus (1 - \alpha) \delta_G \sim \delta_{\alpha F + (1 - \alpha) G}$$

Next property imposes ITRU only with respect to lotteries involving degenerate menus, we name it Degenerate ITRU, more precisely:

(DITRU). For any $f, g \in \mathcal{F}$ and $\alpha \in (0, 1)$,

$$\alpha \delta_f \oplus (1 - \alpha) \delta_g \sim \delta_{\alpha f + (1 - \alpha) g}$$

Before stating the main theorem, we formally introduce the Costly Information Acquisition (CIA) and the Subjective Learning (SL) representations of preferences. First, we need some preliminary definitions. The individual has a prior $\hat{p} \in \Delta(\Omega)$ representing her belief. Information acquisition is modeled through the observation of a noisy signal and the formation of a posterior $p \in \Delta(\Omega)$. Each signal induces a distribution over posteriors $\pi \in \Delta(\Delta(\Omega))$, with the property that the expected posterior equals the prior. We identify a signal with the distribution it induces over posteriors. Hence, for a given prior \hat{p} , the convex set of signals is given by the following:

$$\Pi(\hat{p}) = \left\{ \pi \in \Delta(\Delta(\Omega)) : \int_{\Delta(\Omega)} p \pi(dp) = \hat{p} \right\}$$

The next definition introduces the information cost function:

Definition 1. Given a prior $\hat{p} \in \Delta(\Omega)$, a function $c : \Pi(\hat{p}) \rightarrow [0, \infty]$ is an information cost function if it is lower semicontinuous and it satisfies:

- (i) $c(\pi) = 0$ whenever $\pi(\hat{p}) = 1$.
- (ii) $c(\alpha\pi + (1 - \alpha)\rho) \leq \alpha c(\pi) + (1 - \alpha)c(\rho)$ for all $\pi, \rho \in \Pi(\hat{p})$ and $\alpha \in (0, 1)$.
- (iii) $c(\rho) \leq c(\pi)$ for all $\pi, \rho \in \Pi(\hat{p})$ and $\int_{\Delta(\Omega)} \sigma(p) \pi(dp) \geq \int_{\Delta(\Omega)} \sigma(p) \rho(p)$ for all convex and continuous $\sigma : \Delta(\Omega) \rightarrow \mathbb{R}$.

Property (i) states that acquiring no information is costless. Property (ii) and lower semicontinuity are regularity conditions satisfied by most cost functions used in the literature (e.g. Shannon Information). Condition (iii) states that more informative signals (in the sense of Blackwell) are more costly.

Definition 2. A binary relation \succsim has a Costly Information Acquisition representation if, there exists a tuple $(V, \Pi(\hat{p}), \hat{p}, u, c)$, where $\Pi(\hat{p}) \subset \Delta(\Delta(\Omega))$ is a set of signals, $\hat{p} \in \Delta(\Omega)$ is a prior, $u : X \rightarrow \mathbb{R}$ is a Bernoulli utility and $c : \Delta(\Delta(\Omega)) \rightarrow [0, \infty]$ is an information cost function and $V : \mathcal{A} \rightarrow \mathbb{R}$ is such that, for all $P, Q \in \Delta(\mathcal{A})$, $P \succsim Q \iff E_P[V] \geq E_Q[V]$ and

$$V(F) = \max_{\pi \in \Pi(\hat{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u \circ f \, dp \right) d\pi - c(\pi) \right]$$

The Subjective Learning representation of [Dillenberger et al. \(2014\)](#) is a particular case of the CIA representation in which information is exogenously fixed. Formally,

Definition 3. A binary relation \succsim has a Subjective Learning representation if, there exists a tuple (V, π, u) where, $\pi \in \Delta(\Delta(\Omega))$ is a signal, $u : X \rightarrow \mathbb{R}$ is a Bernoulli utility and $V : \mathcal{A} \rightarrow \mathbb{R}$ is such that, for all $P, Q \in \Delta(\mathcal{A})$, $P \succsim Q \iff E_P[V] \geq E_Q[V]$ and

$$V(F) = \int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u \circ f \, dp \right) d\pi$$

In the Subjective Learning model, the choice of information is not part of the decision process.

The next theorem is the main result of the paper:

Theorem 1. Given a binary relation \succsim on $\Delta(\mathcal{A})$:

1. \succsim satisfies Preference, Ex-ante Independence, PERU and DITRU, if and only if, \succsim has a Costly Information Acquisition representation.
2. \succsim satisfies Preference, Ex-ante Independence and ITRU, if and only if, \succsim has a Subjective Learning representation.

Theorem 1 characterizes the relation between attitude toward the timing of resolution of uncertainty and the costly acquisition of information. A preference for early resolution, which becomes indifference when restricted to de-

generate menus, characterizes the costly information acquisition model.⁷ Indeed, when mixing singletons, the timing of resolution of uncertainty is ineffective. This is a natural consequence of the information acquisition problem modelled by CIA. When facing two degenerate menus of acts, information is not valuable, since information is instrumental to perform an optimal choice *from* the menu. For singleton menus there is no gain in acquiring information. As a consequence, costly information acquisition does not play a role in this situation. This is the main difference between the CIA representation and a pure preference for early resolution of uncertainty.

When the timing of resolution of uncertainty is not relevant (e.g. ITRU), the preference is represented by the Subjective Learning model. The individual acquires information observing the exogenous signal π .

To highlight the strength of PERU we provide a different axiomatization of the CIA model that is more directly comparable with that of [De Oliveira et al. \(2014\)](#). Indeed, although PERU can be compared with Axiom 6 (Aversion to Randomization) in [De Oliveira et al. \(2014\)](#), DITRU is difficult to compare to Axiom 4 (Weak Singleton Independence).⁸ In our alternative axiomatization, we assume PERU and we relax Axiom 4 (WSI) to Weak Degenerate Independence. The latter imposes independence for ex post mixtures when restricted to singletons hence, it is strictly weaker than WSI (see also Section 3.2):

Axiom (Weak Degenerate Independence (WDI)). *For any $f, g, h, h' \in \mathcal{F}$ and $\alpha \in (0, 1)$,*

$$\delta_{\alpha f + (1-\alpha)h} \succ \delta_{\alpha g + (1-\alpha)h} \implies \delta_{\alpha f + (1-\alpha)h'} \succ \delta_{\alpha g + (1-\alpha)h'}$$

Then, we have the following corollary:

Corollary 1. *A binary relation \succ satisfies axioms Preference, Ex-ante Independence, WDI and PERU, if and only if, \succ has a Costly Information Acquisition representation.*

It follows from the previous corollary that the weakening of Axiom 4 (WSI) of [De Oliveira et al. \(2014\)](#) to the WDI axiom can be compensated by the additional strength of PERU, with respect to Axiom 6 (Aversion to Randomization)

⁷More precisely, the CIA model is characterized by PERU with, at least, one strict preference.

⁸WSI (Axiom 4) postulates that, for all menus F, G , all $\alpha \in (0, 1)$ and all acts h, h' , $\alpha F + (1 - \alpha)h \succ \alpha G + (1 - \alpha)h \implies \alpha F + (1 - \alpha)h' \succ \alpha G + (1 - \alpha)h'$. In our setting it would be: for all $F, G \in \mathcal{A}$ and $h, h' \in \mathcal{F}$ and all $\alpha \in (0, 1)$, $\delta_{\alpha F + (1-\alpha)h} \succ \delta_{\alpha G + (1-\alpha)h} \implies \delta_{\alpha F + (1-\alpha)h'} \succ \delta_{\alpha G + (1-\alpha)h'}$.

in [De Oliveira et al. \(2014\)](#). In other words, we *derive* from PERU, WDI and Ex-ante Independence their Axiom 4 (WSI), rather than assuming it.

3.1. Optimal information and ITRU

In this section we show that the optimal signals associated to a given menu, i.e. the ones solving the optimization problem in (CIA), convey information about the attitude toward the resolution of uncertainty. In particular, if two menus share an optimizing signal, then the individual is indifferent to mixing them before or after acquiring information. More precisely, consider the set of optimal signals for a given menu $F \in \mathcal{A}$,

$$\partial V(F) = \operatorname{argmax}_{\pi \in \Pi(\hat{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u \circ f \, dp \right) d\pi - c(\pi) \right]$$

and take two menus $F, G \in \mathcal{A}$. If there exists a signal π^* belonging to $\partial V(F)$ and to $\partial V(G)$, it means that the optimal information acquisition related to F and the one related to G have, at least, a common solution. The mere existence of such signal is sufficient to impose indifference toward the timing of randomization, when it involves F and G . Intuitively, if π^* is optimal for both F and G , the individual can always acquire π^* , so she is indifferent to the timing of the randomization, as in the SL model. More precisely:

Proposition 1. *Suppose that \succsim has a CIA representation, then the following are equivalent:*

1. $\partial V(F) \cap \partial V(G) \neq \emptyset$
2. $\alpha \delta_F \oplus (1 - \alpha) \delta_G \sim \delta_{\alpha F + (1 - \alpha) G}$ for all $\alpha \in (0, 1)$.

The result *does not* imply that the individual will necessarily acquire π^* when evaluating F and G . However, the existence of a common optimal signal is sufficient to guarantee indifference. After all, the individual can always acquire π^* . The opposite of condition 2. in Proposition 1, has been interpreted in [De Oliveira et al. \(2014\)](#) as a "reallocation of attention between F and G ", that is, the tendency to focus on different sources of informations depending on their relevance when facing different menus. This can be used to extend their comparative static analysis. Suppose we can observe the preferences \succsim_1 and \succsim_2 of two individuals and assume V_1 and V_2 are their representations. The following definition can be found in [De Oliveira et al. \(2014\)](#):

Definition 4. V_1 has a greater tendency to reallocate attention than V_2 if, for all menus $F, G \in \mathcal{A}$, $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ implies $\partial V_1(F) \cap \partial V_1(G) = \emptyset$.

The following is a comparative notion of preference for early resolution of uncertainty.

Definition 5. V_1 has a greater preference for early resolution of uncertainty than V_2 if, for all menus $F, G \in \mathcal{A}$:

$$\alpha \delta_F \oplus (1 - \alpha) \delta_G \succ_2 \delta_{\alpha F + (1 - \alpha) G} \implies \alpha \delta_F \oplus (1 - \alpha) \delta_G \succ_1 \delta_{\alpha F + (1 - \alpha) G}$$

Whenever individual 2 strictly prefers early resolution of uncertainty, so does individual 1. The following theorem establishes an equivalence between greater tendency to relocate attention and greater preference for early resolution of uncertainty.

Theorem 2. *The following are equivalent:*

1. V_1 has a greater preference for early resolution of uncertainty than V_2 .
2. V_1 has a greater tendency to reallocate attention than V_2 .

The two conditions are equivalent, indeed, as proved in [De Oliveira et al. \(2014\)](#), a tendency to reallocate attention is related to a form of preference for early resolution of uncertainty (as defined in their condition (ii) of Proposition 3).

3.2. Relation with [De Oliveira et al. \(2014\)](#)

[De Oliveira et al. \(2014\)](#) proposed an axiomatization of the CIA model that is the starting point of the present work. There are, however, some substantial differences. First, we based our axiomatic on the distinction between early and late resolution of ex-ante uncertainty. Indeed, our setting allows to distinguish between randomizations taking place before or after information is acquired. The main axioms in our representation are PERU and DITRU. They capture the attitude toward the timing of resolution of uncertainty for a CIA individual. In their work, [De Oliveira et al. \(2014\)](#) introduced a weak form of preference for early resolution of uncertainty, Axiom 6 (Aversion to Randomization). It postulates that, if $F \sim G$, then $F \succ \alpha F + (1 - \alpha)G$, in our setting this would be equivalent to $\delta_F \sim \delta_G$ implies $\delta_F \succ \delta_{\alpha F + (1 - \alpha)G}$. When the agent is indifferent between

two menus, she may strictly prefer to commit to one menu rather than randomizing. Although this is a weak form⁹ of preference for "early resolution of uncertainty", in their setting there is no explicit modeling of the timing of resolution of uncertainty since, the randomization takes place always after information is acquired (attention is allocated). Moreover, the previous axiom is sufficient to characterize the CIA representation only when coupled with their Axiom 4 (Weak Singleton Independence). To see that Aversion to Randomization does not characterize CIA, consider the following example:

Example 1. Let $c_i(\pi) : \Delta(\Delta(\Omega)) \rightarrow [0, \infty]$ be information costs, for $i = 1, 2$, such that $\max_{\pi \in \Pi(\hat{p})} \left[\frac{1}{c_i(\pi)} \right] = 1$ for $i = 1, 2$. Define:

$$V(F) = \max_{\pi \in \Pi(\hat{p})} \left[\frac{\left(\int \max_{f \in F} \left(\int u(f) dp \right) \pi(dp) \right)^+}{c_1(\pi)} - \frac{\left(\int \max_{f \in F} \left(\int u(f) dp \right) \pi(dp) \right)^-}{c_2(\pi)} \right]$$

where $(\cdot)^+$ and $(\cdot)^-$ denote the positive and negative part. V has the following properties (see Section [Appendix B.1](#)): it satisfies Axioms 6 (AR) in [De Oliveira et al. \(2014\)](#), it satisfies DITRU, it violates Axiom 4 (WSI) in [De Oliveira et al. \(2014\)](#) and it violates PERU¹⁰.

Example 1 provides a representation of preferences that is not a CIA model (since the cost of information is multiplicative, hence it violates Axiom 4 (WSI) in [De Oliveira et al. \(2014\)](#)) and satisfies DITRU (and WDI) and Axiom 6 (AR). To the contrary, it does not satisfy PERU. Assuming PERU in Example 1 would imply the existence of an alternative representation of V of the CIA form. Example 1 shows that, even when DITRU (or WDI) holds, Axiom 6 (AR) alone may not be sufficient to characterize CIA. The additional power of Axiom 4 (WSI) is needed. This is clear from our second axiomatization (Corollary 1), where we show the sufficiency of PERU (and Ex-ante Independence) in characterizing the CIA model. We assumed Weak Degenerate Independence, rather than Axiom 4 (WSI) in [De Oliveira et al. \(2014\)](#) and PERU, rather than Axiom 6 (AR) in [De Oliveira et al. \(2014\)](#) and we derive the CIA representation. In other words, the strength of PERU can compensate the weakening of WSI to WDI. Summing

⁹To see that Ex-ante Independence and PERU imply Aversion to Randomization, suppose $\delta_F \sim \delta_G$, by Ex-ante independence, $\alpha \delta_F \oplus (1 - \alpha) \delta_G \sim \delta_G$, by PERU, $\delta_G \succsim \delta_{\alpha F + (1 - \alpha) G}$.

¹⁰The functional V is quasi-convex and positively homogeneous. It is convex (hence it would satisfy PERU), if and only if, $c_1 \leq c_2$ (see ([Cerreià-Vioglio et al., 2011a](#), Prop. 6).)

up, in our richer setting and given Axiom Preference and Ex-ante Independence, the CIA model can be essentially characterized by one of the following pairs of axioms: Axiom 6 (Aversion to Randomization) and Axiom 4 (Weak Singleton Independence), as in [De Oliveira et al. \(2014\)](#), PERU and DITRU (Theorem 1), PERU and Weak Degenerate Independence (Corollary 1).

To conclude, our axiomatizations constitute an alternative to the one proposed in [De Oliveira et al. \(2014\)](#). We focus on the effect that the timing of resolution of uncertainty has on preferences when information is costly. It is meant to link two different branches of literature and to give a complete analysis of the relation between costly information acquisition and the preference for early resolution of uncertainty.

3.3. Relation with [Ergin and Sarver \(2015\)](#)

The approach we use in this paper is related to a recent work of [Ergin and Sarver \(2015\)](#). They studied preferences toward the timing of resolution of uncertainty when the objects of choice are lotteries over menus of lotteries. Their representation is formally similar to the one in Definition 2 and it ranks lotteries over menus of lotteries according to $P \mapsto E_P[V]$ where:

$$V(F) = \max_{\theta \in \Theta} \int_S \max_{\beta \in F} u(\beta, s; \theta) \pi(ds, \theta) - c(\theta)$$

where Θ is interpreted as a set of hidden actions a decision maker could take before the resolution of uncertainty. They also impose additional axioms to specialize the previous representation to the Costly Contemplation representation introduced in [Ergin and Sarver \(2010a\)](#). In that case, $\theta \in \Theta$ are "contemplation strategies", i.e. observations of a signal about the state space. This last representation is the closest to the one studied in this work. There are, however, some important differences: first, (ex post) uncertainty in [Ergin and Sarver \(2015\)](#) is subjective. In their work, menus contain objective lotteries and the probabilities $\pi(\cdot, \theta)$ are defined over future von Neumann-Morgenstern preferences, as typically studied in preference over menus of lotteries. In this work, we assume the existence of an objective state space Ω and a decision maker who acquires information to form a posterior over Ω and choose from menus of Anscombe-Aumann acts: functions from the objective state space to consequences. Therefore, information is used to update a prior over an "exogenous" state space Ω , rather than updating a prior over an "endogenous" space S of future tastes. In the present work, as well as in that of [De Oliveira et al. \(2014\)](#), tastes are fixed (u is unique) and the preference for flexibility, postulated by the

Dominance axiom, follows from possible gains due to information acquisition (see also the discussion in [De Oliveira et al. \(2014\)](#)).

The second main distinction concerns the separation of tastes and beliefs. We derive a representation that separates tastes and beliefs, something that is not possible in the setting of [Ergin and Sarver \(2015\)](#). Their Costly Contemplation identifies a unique set of measures that they interpret as contemplation strategies. However, the separate identification of utilities and probabilities cannot be achieved, due to state-dependence of the utilities $u(\cdot, s; \theta)$ (see [Ergin and Sarver, 2015](#), Pag. 517).

In addition, the same discussion on the role of Blackwell order contained in [De Oliveira et al. \(2014\)](#) applies here. In the CIA model, the Blackwell order has a central role in the identification of the cost function ([De Oliveira et al., 2014](#), Section 3.3 in) and in the comparative static analysis of "attentiveness" (Section 3.4 in [De Oliveira et al., 2014](#)). The use of Blackwell order has no counterpart in the work of [Ergin and Sarver \(2015\)](#).

AppendixA. Preliminaries

For completeness, we introduce some notation taken from [De Oliveira et al. \(2014\)](#). $C(\Delta(\Omega))$ is the linear space of continuous real-valued function defined on $\Delta(\Omega)$. $ca(\Delta(\Omega))$ is the linear space of signed measures of bounded variation defined on $\Delta(\Omega)$. $C(\Delta(\Omega))$ is equipped with the supnorm and $ca(\Delta)$ with the weak* topology. $ca(\Delta(\Omega))$ is the continuous dual of $C(\Delta(\Omega))$. We denote by Φ the set of convex functions in $C(\Delta(\Omega))$, it is a closed convex cone containing the zero function.

A functional $W : C(\Delta(\Omega)) \rightarrow \mathbb{R}$ is said to be normalized if $W(k) = k$ for all $k \in \mathbb{R}$ (where we identify the constant function $1_{\Delta(\Omega)}k$ with k); monotone, if $W(a) \geq W(b)$ whenever $a(p) \geq b(p)$ for all $p \in \Delta(\Omega)$; translation invariant if $W(a+k) = W(a) + k$ for all $k \in \mathbb{R}$. A translation invariant functional is (Lipschitz) continuous.

The next proposition will be useful in the proof of the main theorem. Let $\Gamma \subseteq C(\Delta(\Omega))$, such that Γ is convex and $\Gamma = \Gamma + \mathbb{R}$.

Proposition 2. *A convex, continuous and normalized functionals $W : \Gamma \rightarrow \mathbb{R}$ is translation invariant.*

Proof. Of Proposition 2. By convexity and normalization, for $\sigma \in \Gamma$ and $k \in \mathbb{R}$,

$$\begin{aligned} W(\sigma + k) &= W\left(\alpha \frac{\sigma}{\alpha} + (1 - \alpha) \frac{k}{(1 - \alpha)}\right) \\ &\leq \alpha W\left(\frac{\sigma}{\alpha}\right) + (1 - \alpha) W\left(\frac{k}{1 - \alpha}\right) \\ &= \alpha W\left(\frac{\sigma}{\alpha}\right) + (1 - \alpha) \frac{k}{1 - \alpha} \end{aligned}$$

since it is true for all $\alpha \in (0, 1)$, by continuity, it is true as $\alpha \rightarrow 1$, then,

$$W(\sigma + k) \leq W(\sigma) + k$$

For any $\sigma \in \Gamma$ and $k \in \mathbb{R}$,

$$\begin{aligned} W(\sigma + k) &\leq W(\sigma) + k = W(\sigma + k - k) + k \\ &\leq W(\sigma + k) - k + k = W(\sigma + k) \end{aligned}$$

Then $W(\sigma + k) = W(\sigma) + k$ for all $\sigma \in \Gamma$ and $k \in \mathbb{R}$. □

Given an affine utility $u : X \rightarrow \mathbb{R}$, we denote by $\sigma_F : \Delta(\Omega) \rightarrow \mathbb{R}$, the function

$$\sigma_F(p) = \max_{f \in F} \int_{\Omega} u \circ f dp \quad \forall p \in \Delta(\Omega) \quad (\text{A.1})$$

for some finite menu F . The set of such functions is denoted by $\hat{\Phi}$. If $u(X) = \mathbb{R}$, $\hat{\Phi}$ has the following properties (De Oliveira et al., 2014).

1. $\hat{\Phi} + \mathbb{R} = \hat{\Phi}$.
2. $\hat{\Phi} + \mathbb{R}$ is dense in Φ .

Appendix B. Proofs

By the Mixture Space theorem (Herstein and Milnor, 1953), axioms Weak Order, Continuity and Ex-ante Independence are necessary and sufficient to the existence of a continuous function $V : \mathcal{A} \rightarrow \mathbb{R}$ such that

$$P \succsim Q \iff E_P[V] \geq E_Q[V]$$

Proof. Of Theorem 1. Part 1. The fact that a preference with a CIA representation implies the axioms is straightforward. To prove the converse implication of Theorem 1 we need a series of preliminary results.

The next lemma introduces a representation of \succsim over degenerate menus:

Lemma 1. *There exists an affine utility $u : X \rightarrow \mathbb{R}$ with unbounded (above and below) range and a probability $\hat{p} \in \Delta(\Omega)$ such that*

$$V(f) = \int_{\Omega} u(f(\omega)) \hat{p}(d\omega)$$

represents \succsim over $\Delta(\mathcal{F})$. Moreover, \hat{p} is unique and u is unique up to positive affine transformations.

Proof. Proof of Lemma 1. The result follows from (Maccheroni et al., 2006, Corollary 20), once we prove that $\delta_f \sim \delta_g$ implies $\delta_{\gamma f + (1-\gamma)g} \sim \delta_f$ for all $\gamma \in (0, 1)$. Suppose $\delta_f \sim \delta_g$, by Ex-ante Independence, for all $H \in \mathcal{A}$ and $\gamma \in (0, 1)$, $\gamma \delta_f \oplus (1-\gamma) \delta_H \sim \gamma \delta_g \oplus (1-\gamma) \delta_H$. Considering all the singletons $\gamma \delta_f \oplus (1-\gamma) \delta_h \sim \gamma \delta_g \oplus (1-\gamma) \delta_h$ then, by DITRU, $\delta_{\gamma f + (1-\gamma)h} \sim \delta_{\gamma g + (1-\gamma)h}$. Hence \succsim satisfies the condition of (Maccheroni et al., 2006, Corollary 20). The uniqueness part and unboundedness above and below of the function u follow from standard arguments (see Cerreia-Vioglio et al., 2011b). □

Lemma 2. *For each $F \in \mathcal{A}$, there exists a $x_F \in X$ such that $\delta_F \sim \delta_{x_F}$.*

Proof. Of Lemma 2. Since Ω and F are finite, there exist a best outcome $b \in X$ and a worst outcome $w \in X$ such that $\delta_b \succ \delta_F \succ \delta_w$. By (Herstein and Milnor, 1953, Th. 1) there exists $\hat{\gamma} \in [0, 1]$ such that $\hat{\gamma}\delta_b \oplus (1 - \hat{\gamma})\delta_w \sim \delta_F$. By DITRU $\delta_{\hat{\gamma}b + (1 - \hat{\gamma})w} \sim \delta_F$. □

Now, let's define the functional $W : \hat{\Phi} \rightarrow \mathbb{R}$ as $W(\sigma_F) \triangleq V(x_F)$ with $x_F \in X$ and $\delta_{x_F} \sim \delta_F$ and σ_F is given by Eq. (A.1). To see that W is uniquely defined, assume $\delta_F \sim \delta_{x_F} \sim \delta_{y_F}$, then $W(\sigma_F) = V(x_F) = V(y_F) = W(\sigma_F)$. To conclude we need to show that $\sigma_F = \sigma_G$ implies $\delta_F \sim \delta_G$. To see this, notice that Claim 1, Claim 4 and Claim 5 of De Oliveira et al. (2014) hold in our setting. Then, W is well-defined.

Lemma 3. *W is monotone (with respect to point-wise order) and normalized. It is convex if and only if \succ satisfies PERU.*

Proof. Of Lemma 3. Monotonicity follows from the definition of W . The constant functions in $\hat{\Phi}$ are those taking values in $u(X) = \mathbb{R}$. To see normalization, take $x \in X$, then $W(\sigma_x) = u(x) = \sigma_x$ by definition. To see convexity, for all $\sigma_F, \sigma_G \in \hat{\Phi}$ and all $\alpha \in (0, 1)$, PERU implies

$$\begin{aligned} \alpha\delta_F \oplus (1 - \alpha)\delta_G &\succ \delta_{\alpha F + (1 - \alpha)G} \iff \\ \alpha V(F) + (1 - \alpha)V(G) &\geq V(\alpha F + (1 - \alpha)G) \iff \\ \alpha W(\sigma_F) + (1 - \alpha)W(\sigma_G) &\geq W(\sigma_{\alpha F + (1 - \alpha)G}) = W(\alpha\sigma_F + (1 - \alpha)\sigma_G) \end{aligned}$$

□

Since W is convex, continuous and normalized, by Proposition 2, W is also translation invariant.

To sum up, W is a monotone, normalized, convex and translation invariant functional, by Cerreia-Vioglio et al. (2014, Prop. 2) it is a niveloid. By the properties in section AppendixA and Claim 7 and 8 of De Oliveira et al. (2014), W can be rewritten as

$$W(\sigma_F) = \max_{\pi \in \Pi(\hat{p})} \int_{\Delta(\Omega)} \max_{f \in F} \left(\int u \circ f dp \right) \pi(dp) - c(\pi)$$

where $c : \Pi(\hat{p}) \rightarrow [0, \infty]$ is given by

$$c(\pi) = \sup_{F \in \mathcal{A}} \left[\left(\int u \circ f dp \right) \pi(dp) - W(\sigma_F) \right]$$

This concludes the proof of part a).

For part 2., the fact that a SL representation implies the axioms is trivial. For the converse direction, ITRU clearly implies DITRU, hence Lemma 1 holds. We can define W and V as in the proof of Theorem 1. To prove the result we need to show that our axioms imply Axiom 9 in De Oliveira et al. (2014). Suppose that $\delta_F \sim \delta_G$, by Ex-ante Independence, $\gamma\delta_F \oplus (1-\gamma)\delta_H \sim \gamma\delta_G \oplus (1-\gamma)\delta_H$ for all $H \in \mathcal{A}$. By ITRU, $\delta_{\gamma F + (1-\gamma)H} \sim \delta_{\gamma G + (1-\gamma)H}$. Taking $H = G$, we have $\delta_{\gamma F + (1-\gamma)G} \sim \delta_G$. Applying (De Oliveira et al., 2014, Prop. 7) we have the result. \square

Proof. Of Corollary 1. The fact that CIA implies the axioms is straightforward. To prove the other implication, we follow the proof of Theorem 1. First, we prove that WDI, together with the other axioms, implies the independence axiom over degenerate menus. Suppose $\delta_f \sim \delta_g$, by PERU and Ex-ante Independence, $\delta_f \succ \delta_{\alpha f + (1-\alpha)g}$. By WDI, $\delta_{\alpha f + (1-\alpha)f} \succ \delta_{\alpha f + (1-\alpha)g}$ implies $\delta_{\alpha g + (1-\alpha)f} \succ \delta_{\alpha g + (1-\alpha)g} = \delta_g$. By PERU and Ex-ante Independence again, $\delta_g \succ \delta_{\alpha g + (1-\alpha)f}$, hence $\delta_f \sim \delta_g \sim \delta_{\alpha g + (1-\alpha)f}$. Therefore, Lemma 1 holds. To prove that each $F \in \mathcal{A}$ has a certainty equivalent, notice that finiteness of F and of Ω imply the existence of a best $b \in X$ and worst outcome $w \in X$ such that $\delta_b \succ \delta_F \succ \delta_w$. By Continuity, there exists $\hat{\alpha} \in [0, 1]$, such that $\delta_F \sim \hat{\alpha}\delta_b \oplus (1-\hat{\alpha})\delta_w$. By PERU, $\delta_F \sim \hat{\alpha}\delta_b \oplus (1-\hat{\alpha})\delta_w \succ \delta_{\hat{\alpha}b + (1-\hat{\alpha})w}$. Suppose, by contradiction, that $\hat{\alpha}\delta_b \oplus (1-\hat{\alpha})\delta_w \succ \delta_{\hat{\alpha}b + (1-\hat{\alpha})w}$, then $\hat{\alpha}V(b) + (1-\hat{\alpha})V(w) > V(\hat{\alpha}b + (1-\hat{\alpha})w)$. By Lemma 1, $\hat{\alpha}u(b) + (1-\hat{\alpha})u(w) > u(\hat{\alpha}b + (1-\hat{\alpha})w)$, contradicting affinity of u . Hence, $\delta_F \sim \hat{\alpha}\delta_b \oplus (1-\hat{\alpha})\delta_w \sim \delta_{\hat{\alpha}b + (1-\hat{\alpha})w}$.

The rest of the proof is identical to the one of Theorem 1. \square

Before proving Proposition 1, notice that $\partial V(F)$ is equal to the subdifferential of $W(\sigma_G)$ (see De Oliveira et al., 2014).

Proof. Of Proposition 1. (1.) implies (2.). Assume $\pi^* \in \partial W(\sigma_F) \cap \partial W(\sigma_G) \neq \emptyset$, by the definition of subdifferential

$$W(\sigma) \geq W(\sigma_F) + \langle \sigma - \sigma_F, \pi^* \rangle \text{ and } W(\sigma) \geq W(\sigma_G) + \langle \sigma - \sigma_G, \pi^* \rangle \quad (\text{B.1})$$

for all $\sigma \in C(\Delta(\Omega))$. Then, choosing $\sigma = \sigma_G$ in the first inequality and $\sigma = \sigma_F$ in the second gives:

$$W(\sigma_F) - W(\sigma_G) = \langle \sigma_F - \sigma_G, \pi^* \rangle \quad (\text{B.2})$$

Now, take $\sigma = \alpha\sigma_F + (1 - \alpha)\sigma_G$ for some $\alpha \in (0, 1)$ and plug it into the right-hand side of Eq. (B.1), then

$$\begin{aligned} W(\alpha\sigma_F + (1 - \alpha)\sigma_G) &\geq W(\sigma_G) + \alpha\langle\sigma_F - \sigma_G, \pi^*\rangle \\ &= W(\sigma_G) + \alpha(W(\sigma_F) - W(\sigma_G)) \\ &= \alpha W(\sigma_F) + (1 - \alpha)W(\sigma_G) \end{aligned}$$

where the first equality follows from Eq. (B.2). The reverse inequality follows from convexity of W .

(2.) implies (1.) Assume 2. and, by contrapositive, $\partial W(\sigma_F) \cap \partial W(\sigma_G) = \emptyset$. Take $\pi \in \partial W(\alpha\sigma_F + (1 - \alpha)\sigma_G)$, it is non-empty by Ergin and Sarver (2010b, Lemma 2.5), then $\pi \notin \partial W(\sigma_F)$ or $\pi \notin \partial W(\sigma_G)$ by the initial assumption, so $W(\sigma_F) \geq \langle\sigma_F, \pi\rangle - W^*(\pi)$ and $W(\sigma_G) \geq \langle\sigma_G, \pi\rangle - W^*(\pi)$ and one of these inequalities must be strict. Then, for all $\alpha \in (0, 1)$:

$$\begin{aligned} \alpha W(\sigma_F) + (1 - \alpha)W(\sigma_G) &> \alpha\langle\sigma_F, \pi\rangle + (1 - \alpha)\langle\sigma_G, \pi\rangle - W^*(\pi) \\ &= \langle\alpha\sigma_F + (1 - \alpha)\sigma_G, \pi\rangle - W^*(\pi) \\ &= W(\alpha\sigma_F + (1 - \alpha)\sigma_G) \end{aligned}$$

a contradiction to (2.) and where the last equality follows from $\pi \in \partial W(\alpha\sigma_F + (1 - \alpha)\sigma_G)$. \square

Proof. Of Theorem 2. 1. implies 2. By contrapositive, assume 1. and the existence of $F, G \in \mathcal{A}$ such that $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ and $\partial V_1(F) \cap \partial V_1(G) \neq \emptyset$. By Prop. 1 and PERU, $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ implies $\alpha\delta_F \oplus (1 - \alpha)\delta_G \succ_2 \delta_{\alpha F + (1 - \alpha)G}$. By Prop. 1, $\partial V_1(F) \cap \partial V_1(G) \neq \emptyset$ implies $\alpha V_1(F) + (1 - \alpha)V_1(G) = V_1(\alpha F + (1 - \alpha)G)$, a contradiction to 1.

2. implies 1. By contrapositive, assume 2. and the existence of $F, G \in \mathcal{A}$ and $\alpha' \in (0, 1)$ such that $\alpha'\delta_F \oplus (1 - \alpha')\delta_G \succ_2 \delta_{\alpha'F + (1 - \alpha')G}$ and $\delta_{\alpha'F + (1 - \alpha')G} \succsim_1 \alpha'\delta_F \oplus (1 - \alpha')\delta_G$. By PERU, the last becomes $\alpha'\delta_F \oplus (1 - \alpha')\delta_G \sim_1 \delta_{\alpha'F + (1 - \alpha')G}$. This is equivalent to $\alpha'V_1(F) + (1 - \alpha')V_1(G) = V_1(\alpha'F + (1 - \alpha')G)$, by Prop. 1, the last equality must be true for all $\alpha \in (0, 1)$. Therefore, $\alpha'\delta_F \oplus (1 - \alpha')\delta_G \sim_1 \delta_{\alpha'F + (1 - \alpha')G}$ implies $\partial V_1(F) \cap \partial V_1(G) \neq \emptyset$. However, $\alpha'\delta_F \oplus (1 - \alpha')\delta_G \succ_2 \delta_{\alpha'F + (1 - \alpha')G}$ implies $\partial V_2(F) \cap \partial V_2(G) = \emptyset$, a contradiction to 2. \square

Appendix B.1. Calculations for Example 1.

From the analysis of positively homogeneous functionals of Cerreia-Vioglio et al. (2011a), it is immediate to see that $V(F)$, seen as a functional $W : C(\Delta(\Omega)) \rightarrow$

\mathbb{R} , is quasi-convex (hence it satisfies AR), positively homogeneous and it is not translation invariant. It is convex (hence it would satisfy PERU), if and only if, $c_1 \leq c_2$, see (Cerrei-Vioglio et al., 2011a, Prop. 6). To see DITRU, consider $f, g \in \mathcal{F}$. Then:

$$V(\gamma f + (1 - \gamma)g) = \max_{\pi \in \Pi(\hat{p})} \left[\frac{\left(\left(\int \gamma f u(f) dp + (1 - \gamma) \int u(g) dp \right) \pi(dp) \right)^+}{c_1(\pi)} - \frac{\left(\left(\int \gamma f u(f) dp + (1 - \gamma) \int u(g) dp \right) \pi(dp) \right)^-}{c_2(\pi)} \right]$$

By the property of each signal in $\Pi(\hat{p})$ and the condition on the information cost functions, $V(\gamma f + (1 - \gamma)g) = \gamma \int u(f) d\hat{p} + (1 - \gamma) \int u(g) d\hat{p}$. Which is equal to the value of $\gamma \delta_f + (1 - \gamma) \delta_g$.

References

- Caplin, A. and M. Dean (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, Forthcoming.
- Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011a). Complete monotone quasiconcave duality. *Math. Oper. Res.* 36(2), 321–339.
- Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011b). Uncertainty averse preferences. *Journal of Economic Theory* 146(4), 1275–1330.
- Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2014). Niveloids and their extensions: Risk measures on small domains. *Journal of Mathematical Analysis and Applications* 413(1), 343 – 360.
- De Oliveira, H. (2014). Axiomatic foundations for entropic costs of attention. *Mimeo*.
- De Oliveira, H., T. Denti, M. Mihm, and M. K. Ozbek (2014). Rationally inattentive preferences. *Mimeo*.
- Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014). A theory of subjective learning. *Journal of Economic Theory* 153(0), 287 – 312.
- Ellis, A. (2013). Foundations for optimal inattention. *Mimeo*.

- Ergin, H. and T. Sarver (2010a). A unique costly contemplation representation. *Econometrica* 78(4), 1285–1339.
- Ergin, H. and T. Sarver (2010b). The unique minimal dual representation of a convex function. *Journal of Mathematical Analysis and Applications* 370(2), 600–606.
- Ergin, H. and T. Sarver (2015). Hidden actions and preferences for timing of resolution of uncertainty. *Theoretical Economics* 10(2), 489–541.
- Herstein, I. N. and J. Milnor (1953). An axiomatic approach to measurable utility. *Econometrica, Journal of the Econometric Society*, 291–297.
- Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2009). Rational attention allocation over the business cycle. NBER Working Papers 15450, National Bureau of Economic Research, Inc.
- Kreps, D. M. and E. L. Porteus (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46(1), pp. 185–200.
- Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic Dynamics* 11(2), 366–385.
- Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 74(6), 1447–1498.
- Mackowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review* 99(3), 769–803.
- Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012). Revealed attention. *American Economic Review* 102(5), 2183–2205.
- Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: a new foundation for the multinomial logit model. *American Economic Review* 105(1), 272–98.
- Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory* 145(5), 1837–1864.
- Piermont, E., N. Takeoka, and R. Teper (2015). Learning the krepsian state: exploration through consumption. *Mimeo*.

- Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy* 49(1), 317–356.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50(3), 665–690.
- Tutino, A. (2008). The rigidity of choice: lifetime savings under information-processing constraints. MPRA Paper 16744, University Library of Munich, Germany.
- Van Nieuwerburgh, S. and L. Veldkamp (2009). Information immobility and the home bias puzzle. *Journal of Finance* 64(3), 1187–1215.